Published in final edited form as:
Nat Struct Mol Biol. 2014 February ; 21(2): 198–206. doi:10.1038/nsmb.2764.

Topological Organization of Multi-chromosomal Regions by *Firre*

Ezgi Hacisuleyman[1,2,3,13], **Loyal A. Goff**[2,3,4,13], **Cole Trapnell**[2,3], **Adam Williams**[5], **Jorge Henao-Mejia**[5], **Lei Sun**[6], **Patrick McClanahan**[7], **David G. Hendrickson**[2,3], **Martin Sauvageau**[2,3], **David R. Kelley**[2,3], **Michael Morse**[3], **Jesse Engreitz**[3], **Eric S. Lander**[3], **Mitch Guttman**[8], **Harvey F. Lodish**[6,9,10], **Richard Flavell**[5,11], **Arjun Raj**[7], and **John L. Rinn**[2,3,12,14]

[1]Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts, USA

[2]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts, USA

[3]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts, USA

[4]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[5]Department of Immunobiology, Yale University School of Medicine, New Haven, Connecticut, USA

[6]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, USA

[7]Department of Bioengineering, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[8]Department of Biology, California Institute of Technology, Pasadena, California, USA

[9]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[10]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[11]Howard Hughes Medical Institute, Boston, Massachusetts, USA

[12]Department of Pathology, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA

Abstract

RNA is known to be an abundant and important structural component of the nuclear matrix, including long noncoding RNAs (lncRNA). Yet the molecular identities, functional roles, and localization dynamics of lncRNAs that influence nuclear architecture remain poorly understood. Here, we describe one lncRNA, Firre, that interacts with the nuclear matrix factor hnRNPU, through a 156 bp repeating sequence and Firre localizes across a ~5 Mb domain on the X-

[14]Corresponding author: johnrinn@fas.harvard.edu.
[13]Authors contributed equally to this work.

Accession codes. All sequencing and related data is deposited in the Gene Expression Omnibus, under accession code GSE45157.

Author Contributions

E.H. designed and performed experiments; L.A.G. performed computational analysis; J.L.R. directed research. E.H, L.A.G., J.L.R wrote the manuscript. C.T. and D.R.K. helped with computational analysis. M.S, D.G.H, and P.M., J.E., M.G. A.R. contributed experiments. A.W., J.H. R.F. generated knockout mESCs.

chromosome. We further observed Firre localization across at least five distinct trans-chromosomal loci, which reside in spatial proximity to the *Firre* genomic locus on the X-chromosome. Both genetic deletion of the *Firre* locus or knockdown of hnRNPU resulted in loss of co-localization of these trans-chromosomal interacting loci. Thus, our data suggest a model in which lncRNAs such as Firre can interface with and modulate nuclear architecture across chromosomes.

INTRODUCTION

It has become clear that the mammalian genomes encode many lncRNAs with diverse functions in development and disease[1–6]. Recent work has begun to identify the myriad roles for lncRNAs, including but not limited to forming ribonucleoprotein (RNP) complexes with epigenetic regulatory machinery, transcriptional and post-transcriptional regulation of gene expression, and the formation sub-compartments in the nucleus to mediate higher-order chromosomal architecture[7–14]. Examples of these phenomena can be observed in X-chromosome dosage compensation in mammals. Several lncRNAs have been shown to recruit epigenetic regulatory complexes (e.g. polycomb)[15,16], some of which are brought to the future inactive X-chromosome[17,18]. Specifically, the lncRNA Xist binds to multiple proteins as a “scaffold” to mediate the silencing of genes on the X-chromosome and affect the higher-order chromosomal architecture needed to establish proper epigenetic silencing[14,19,20].

RNA molecules, have long been known to be key constituents of the nuclear matrix[21–24] and several noncoding RNAs[24,25] have been demonstrated to be involved in the formation of nuclear sub-compartments such as the nucleolus and paraspeckles[26] as well as the facilitation of higher-order chromosomal architecture[19,20,27]. However, the diversity of lncRNAs mechanisms, their influence on nuclear architecture and consequent cellular roles remain irresolute.

Here, we identify a long intergenic noncoding RNA (lincRNA) termed Functional Intergenic Repeating RNA Element (Firre) that localizes across a 5 Mb domain around its site of transcription. This domain of Firre localization is also in spatial proximity to at least 5 other trans-chromosomal loci within the nucleus. This cross-chromosomal co-localization requires Firre as genetic deletion of *Firre* results in a loss of spatial proximity between its trans-chromosomal binding sites. We further identified a unique 156 bp repeating RNA domain in the Firre sequence that is required to both interact with the nuclear matrix factor hnRNPU and localize Firre transcripts in a punctate manner in the nucleus. Strikingly, the knockdown of hnRNPU similar to deletion of *Firre* locus results in a loss of spatial proximity between the *Firre* locus and its trans-chromosomal binding sites. Collectively, these findings suggest a model where lncRNAs, such as Firre, can function as nuclear organization factors that interact with and influence higher order nuclear architecture across chromosomes.

RESULTS

Firre is a nuclear-retained and chromatin-associated lncRNA

We previously identified Firre (previously referred to linc-RAP-1) as being required for proper adipogenesis in a loss of function screen in murine adipocyte precursors[28]. A detailed subsequent analysis of Firre revealed many interesting and distinguishing features: (i) a diverse expression pattern of Firre *in vivo*, with enrichment in neural crest tissues as shown by *in situ* hybridization (Fig. 1A-I); (ii) a conserved intergenic human ortholog located on the X-chromosome (Fig. 1J, K); (iii) a unique 156bp Repeating RNA Domain (RRD) that occurs 16 and 8 times in *Mus musculus* (mouse) and *Homo sapien* (human) transcripts,

respectively, with 96% sequence identity within species and 68% across species (Fig. 1J); (iv) numerous alternatively-spliced isoforms with differential inclusion or exclusion of RRD sequences (Supplementary Fig. 1A, B, Supplementary Table 1). (v) Firre transcripts remain stable even after 6 hours of actinomycin D (ActD) treatment (Supplementary Fig. 1C, D).

To further determine the subcellular localization of Firre, we used single molecule RNA fluorescence *in situ* hybridization (FISH) targeting Firre (as described in [29]). We adopted a dual labeling strategy to independently target the introns and exons of Firre, thus marking the site of transcription on the X-chromosome (intronic probes) and the location of the mature transcripts (exonic probes) separately (Fig. 2A, Supplementary Table 2). RNA-FISH analysis revealed an exclusively nuclear and focal distribution for Firre (Fig. 2A, Supplementary Fig. 2) in all cells tested. Notably, Firre exhibits strong expression foci near its site of transcription in both male and female mouse and human ESCs (mESCs, hESCs) (Fig. 2A). We also note that the Firre RNA is localized around its site of transcription, but extend slightly beyond this site in all six human and mouse cell lines tested. Subcellular localization and expression of Firre in cell lines with and without inactive X-chromosomes was similar to that observed in ESCs (Supplementary Fig. 2). Thus, Firre is nuclear-localized and forms expression foci on both X-chromosomes prior to, and after X-chromosome inactivation.

We tested if over-expression of Firre was sufficient to form the observed endogenous local expression foci. Briefly, we ectopically expressed Firre via retrovirus-mediated integration in human and mouse lung fibroblasts, which do not express Firre (Fig. 2B, Supplementary Fig. 3A). We observed many sites of focal accumulation of Firre (Fig. 1B). We obtained similar results upon ectopic expression of Firre in human cells that endogenously express this lncRNA (HEK293 and HeLa) (Supplementary Fig. 3A, B). We repeated the experiments above using isoforms of Firre with (isoform 4; one repeat) or without the RRD (isoform 6). Strikingly, in the absence of RRD, the nuclear localization of Firre is disrupted, and Firre RNA is detected in the cytoplasm (Fig. 2C). Thus, RRD is required for the focal nuclear localization of Firre.

The *Firre* locus escapes X-chromosome inactivation

Our observation from RNA-FISH in female mouse and human ESCs and HEK293s led us to hypothesize that *Firre* might escape X-chromosome inactivation (XCI). To test this, we analyzed the local chromatin environment within the *Firre* locus using existing chromatin immunoprecipitation (ChIP) data for numerous histone modifications and transcription factors. Several of these data are consistent with *Firre* escaping XCI: First, we observed an appreciable depletion of LaminB1 across the mouse *Firre* locus and across the human *FIRRE* locus in various cell lines (Fig. 1K, 2D). LaminB1 is a matrix protein involved in nuclear stability and chromatin organization and is known to mark heterochromatin. The domain of LaminB1 depletion extends precisely across the body of the *Firre* gene but not into the upstream or downstream regions. Second, the *Firre* locus is specifically and significantly ($p<1.0\times10^{-8}$) depleted of trimethylated histone 3 lysine 27 (H3K27me3) in differentiated mESCs and in human cells prior to and after X-chromosome inactivation (Fig. 2D). Third, the *Firre* locus is enriched for trimethylated histone 3 lysine 4 (H3K4me3) with and without Firre transcription (Fig. 2D). Finally, we observed a striking pattern of CCCTC-binding factor (CTCF) (Fig. 2D), which can function as an insulator between chromatin domains and facilitates inter-chromosomal interactions, localization adjacent to almost every exon of Firre[30].

To further test the hypothesis that *Firre* escapes X-chromosome inactivation, we investigated whether Xist RNA itself localizes on the *Firre* locus upon X-chromosome inactivation. Specifically, we examined the localization of Xist on DNA using data

generated by RAP in mouse lung fibroblasts (mLF). In contrast to the enrichment of Xist across most of the X-chromosome[19], we observed a strong and focal depletion in Xist binding at the *Firre* locus; similar to what was observed at genes known to escape XCI (Fig. 2D, bottom track). Interestingly, we note that the Xist-depleted boundaries are consistent with the previously identified boundaries for the lamin-depleted regions. Collectively, these data indicate that the *Firre* locus escapes X-chromosome inactivation and has a notable enrichment for CTCF and H3K4me3 and depletion for H3K27me3 and LaminB1.

Firre localizes to chromatin *in cis* and *trans*

The focal nuclear localization of Firre near its site of transcription led us to identify the direct interactions between Firre and chromatin in the nucleus. To resolve the DNA binding sites of Firre genome-wide, we performed RNA Antisense Purification (RAP)[19]. RAP along with CHiRP[31], CHART[32], ChOP[33] provide genome-wide DNA-binding locations for RNAs by cross-linking chromatin and RNA, followed by the targeting and pull-down of a specific RNA using antisense oligos[19]. We performed RAP in male mESCs using two sets of 120 bp antisense probes targeting Firre and two sets of sense probes as negative controls (Supplementary Table 3), followed by sequencing to identify genomic regions directly bound by Firre. We observed a ~5 Mb domain of Firre localization around the *Firre* locus (Fig. 3A). Strikingly, we also observed five significantly enriched peaks (Cuffdiff2; 1% FDR) of Firre located on chromosomes 2, 9, 15, and 17 that overlap known genes including *Slc25a12*, *Ypel4*, *Eef1a1*, *Atf4*, and *Ppp1r10* (Fig. 3B). Notably, 4 out of 5 of these genes have previously described regulatory roles during adipogenesis[34–37], consistent with our previous study showing the role of Firre in adipogenesis[28]. Expanding this search to regions not overlapping with mRNAs, we observed a total of 34 additional significant (Cuffdiff2; 1% FDR) localization sites for Firre.

Collectively these data suggest that Firre is localized on multiple chromosomes, yet has only one predominant nuclear localization site in male and two in female cells around its site of transcription. These observations suggest two possible models. One possibility is that Firre could be shuttled from its site of transcription to these sites on other chromosomes. Alternatively, the focal localization of Firre to its own genomic locus could serve as a regional organizing factor to bring the trans-interacting sites into the three-dimensional proximity of the *Firre* locus on the X-chromosome.

Firre trans-chromosomal sites are in spatial proximity

In order to determine the nature of the trans-chromosomal interactions for Firre, we performed single molecule RNA co-FISH in mESCs on the transcription sites of Firre and three of the *trans*-interacting genes (*Slc25a12, Ypel4* and *Ppp1r10*) (Fig. 4A). As negative controls, we performed similar RNA co-FISH for Firre and several genes with high expression in mESC that were not detected by RAP as *trans* targets, (e.g. *Pouf51 (Oct4), Nanog* and *Sox2*) (Fig. 4B, Supplementary Fig. 4A). Remarkably, we observed co-localizations between Firre *and* all three trans-sites tested: Slc25a12 (73.9% of cells), Ypel4 (79.4% of cells), and Ppp1r10 (78.1% of cells) (Fig. 4A), and between these trans sites (Fig. 4C, Supplementary Fig. 4B). Conversely, we did not observe any co-localization of Firre and unbound targets Oct4, Nanog, or Sox2 (Fig. 4B, Supplementary Fig. 4A). Thus, these results are consistent with the latter model, where the *Firre* locus resides in three-dimensional proximity to these trans-chromosomal binding sites.

Firre regulates key pluripotency pathways

To determine the functional role of Firre, we generated a Firre knockout male mESC line by deleting the entire *Firre* locus on the X-chromosome (*ΔFirre*). Briefly, we introduced loxP sites in the 5′ and 3′ of the *Firre* locus by a two-step targeting strategy. Then, we infected

the cells that harbor this locus with a Cre plasmid and clonally selected the cells with the proper deletion. Comparison of wild-type and *ΔFirre* growth rates revealed a marked retardation in growth rate and colony formation (Fig. 5A, B). We also note an intermediate growth defect when the cells were grown on a mouse embryonic fibroblast feeder layer (Fig. 5A). The *ΔFirre* cells on feeders were able to form bigger and more colonies in the same amount of time when compared to the *ΔFirre* cells grown without feeders (Fig. 5A).

To identify the gene-pathways and molecular signature that are altered upon deletion of Firre, we conducted massively parallel RNA-sequencing (RNA-Seq) comparisons between wild-type and *ΔFirre*. Briefly, RNA was isolated from three wild-type (WT) and two *ΔFirre* replicate cultures and subjected to paired-end illumina sequencing to a mean depth of ~9×10^6 fragments aligned per replicate. We identified 1077 genes with significant differential expression (Cuffdiff2; 5% FDR) between the WT and *ΔFirre* mESCs (Fig. 5c). Preranked GSEA analysis demonstrated that *ΔFirre* cells were significantly enriched ($p<0.01$) for genes involved in extracellular matrix organization, and cell surface receptor-ligand interactions (Fig. 5d) Conversely, *ΔFirre* mESCs were depleted for genes involved in mRNA processing, nuclear export, and electron transport chain mediated energy metabolism relative to WT (Fig. 5d). Notably, we observed an increase in Tgfβ signaling in the *ΔFirre* mESCs (Fig. 5e). Interestingly Tgfβ signaling is known to be a potent inhibitor of adipogenesis[38], consistent with our previous observation that knockdown of Firre strongly inhibits adipogenesis in mouse preadipocytes[28] and growth defects observed in mESC cultures (Fig. 5a, b).

We next tested whether or not the Firre trans localization sites were affected by the absence of Firre. We did not observe a global enrichment for the five trans-site genes in the list of significantly differentially expressed genes ($p<1.0$; Hypergeometric test). We did observe one key exception, Ppp1r10, one of the 3 validated trans-sites, that was significantly decreased (Cuffdiff2; 1% FDR) in the *ΔFirre* cells relative to WT. However, we cannot preclude the possibility of perturbations to mRNA stability, translation, or processing at any of the remaining trans-sites.

Firre binds hnRNPU in an RRD-dependent manner

We next turned to identify proteins that interact with Firre that might mediate its 5Mb X-chromosome localization. As a first approach to identify the candidate protein partners of Firre, we performed RNA pull-down assays in mouse ESC and adipocyte lysates by biotinylating the RNA, either by body-labeling (*in vitro* transcription) or 3′ end-labeling (pCp-biotin), followed by mass-spectrometry. Unrelated ncRNAs (sense and antisense telomerase RNA TERC) were used as negative controls (Supplementary Fig. 5A). To identify proteins that preferentially co-precipitated with Firre in an RRD-dependent manner, we used five different RRD-positive isoforms and one RRD-negative isoform and took the difference between the peptide counts of the RRD-positive and RRD-negative isoforms (Supplementary Fig. 5B). We repeated the differential analysis for each of the five isoforms and took the top 10% of the differential peptide count scores for each isoform identified in both mESC and adipocyte lysates. Based on the highest unique peptide counts, we identified 8 candidate proteins that physically associate with Firre in an RRD-dependent manner (Supplementary Fig. 5C).

The highest ranked candidate from this analysis was heterogeneous ribonucleoprotein U (hnRNPU) (Supplementary Fig. 5C). hnRNPU was of particular interest because it is required for the proper localization of Xist, interaction with the scaffold attachment regions on DNA, and the formation of highly structured chromatin territories[39–41]. To confirm the interaction between hnRNPU and Firre, we repeated the RNA pull-down experiments described above, and assayed for hnRNPU via Western blotting. In both mESC and

adipocyte lysates, hnRNPU co-precipitated with Firre, but not with the negative controls (Fig. 6A, Supplementary Fig. 5D). We further tested additional hnRNP family members and found either no association or association independent of RRD (Fig. 6A, Supplementary Fig. 5D). Together these data suggest that Firre associates with hnRNPU.

To test if the Firre RRD element is required and sufficient to interact with hnRNPU, we performed RNA pull-downs using single and double copies of the mouse and human RRD sequences in mESC lysates. We observed binding of hnRNPU to both mouse and human synthetic RRD constructs (Fig. 6B, panel 1). Finally, western blot analysis confirmed that hnRNPU binds to the Firre isoform harboring RRD but not to the ΔRRD isoform (Fig. 6B, panel 2). To determine if the hnRNPU–Firre interaction is biologically relevant at endogenous levels, we captured endogenous Firre using complimentary DNA oligos and confirmed that hnRNPU co-purifies specifically with the lncRNA (Fig. 6B, panels 3&4, Supplementary Fig. 6).

To further validate the hnRNPU–Firre interaction, we performed the reciprocal experiment: RNA immunoprecipitation (RIP) targeting hnRNPU. Consistent, with the RNA pull-downs, we found strong enrichment of Firre relative to IgG controls after normalization to total input (Fig. 6C). Finally, analysis of publicly available data from UV-crosslinked RIP targeting hnRNPU verified that Firre directly and specifically binds to hnRNPU (Fig. 6D)[42,43]. Collectively, these results suggest that Firre interacts with hnRNPU, and that the RRD is both required and sufficient for this interaction. \

hnRNPU is required for focal localization of Firre

Based on the requirement for RRD in establishing the proper localization of Firre and its interaction with hnRNPU, we investigated whether hnRNPU, in turn, regulates the spatial expression of Firre. Briefly, we transfected siRNAs targeting hnRNPU in mouse (mESC) and human (HEK293 and HeLa) cell lines and observed a >90% decrease in hnRNPU expression (Supplementary Fig. 6C). We confirmed the previously described role of hnRNPU in Xist localization in HEK293 cells (Supplementary Fig. 6D). Following the knockdown of hnRNPU, we repeated RNA-FISH targeting Firre as above. In all cell lines tested, we observed a strong delocalization of Firre and in several instances even translocation into the cytoplasm (Fig. 6E, F). Thus, both RRD and hnRNPU are required for the proper focal and nuclear localization of Firre and its retention in the nucleus.

hnRNPU is required for proximal trans-localization of Firre

Having found that hnRNPU regulates the specific localization of Firre, we next tested whether hnRNPU is required to maintain the proximal localization of the *Firre* locus and its trans-chromosomal localization sites. To this end, we repeated the RNA co-FISH between Firre and either Slc25a12 or Ypel4 upon siRNA-mediated depletion of hnRNPU in mESCs. In both cases, we observed a considerable decrease in co-localization of each *trans*-site with Firre in the absence of hnRNPU (Fig. 7A).

Firre is required for trans-chromosomal co-localization

To test the functional contribution of Firre to trans-chromosomal co-localization, we repeated the co-FISH experiments between the trans sites in male *ΔFirre* mESCs. Strikingly, the Ppp1r10 and Ypel4 gene loci no longer co-localize in the absence of Firre (Fig. 7B) (15% co-localization in *ΔFirre (panel 2)* relative to 72% in wild-type (panel 1)), thus suggesting a requirement for the *Firre* gene locus facilitating the formation of cross-chromosomal interactions in mESCs. Furthermore, we do not observe co-localization of trans-sites in mLFs that do not express Firre (Fig. 7C, panel 3). Collectively, the above

results suggest a potential role for Firre, along with hnRNPU, in either maintaining or establishing higher-order nuclear architecture.

DISCUSSION

Here, we have identified and characterized a novel ncRNA RNA gene, *Firre*, that has important roles in both cell physiology and nuclear architecture. The numerous properties of Firre shed new insights into: how RNA sequences can result in both *cis* and *trans* localization on chromatin, how these interactions between specific RNA sequences and nuclear matrix factors influence nuclear organization, and how these RNA-protein interaction properties summate to modulate properties of higher order nuclear architecture and the subsequent consequences to cell physiology.

Together these observations propose an intriguing model, in which Firre, and potentially many other lncRNAs[13,22,27,44–47], function as nuclear organization factors (Fig. 4B). Specifically, Firre may serve to interface with and modulate the topological organization of multiple chromosomes[9]. Consistent with this model, genetic deletion of *Firre*, results in a loss of nuclear proximity of several trans-chromosomal loci to the *Firre* locus. Moreover, the proper localization of Firre requires both a specific 156 bp sequence and a physical interaction with hnRNPU to maintain the multi-chromosomal nuclear interactions. Thus, we propose that lncRNAs, through the interaction with nuclear matrix proteins, such as hnRNPU, might impart specificity in organizing a proper “zipcode” of chromosomal territories within the nucleus. For example, either the chromosomal binding of Firre or sequence specific interactions may serve as a cis localization signal in order to initiate the formation of or maintain sub-compartments within the nucleus.

Consistent with our model, two recent studies[19,20] demonstrate that the Xist RNA uses a “local proximity search” to guide its localization across large expanses of the X-chromosome during X-chromosome inactivation. Here, we broaden this phenomenon and show that these interactions are not merely restricted to a single chromosome but extend across multiple chromosomes in regional proximity. Several other observations in this study highlight potential gene regulatory roles of Firre, mediated by trans-chromosomal interactions. Intriguingly, we observe an array of CTCF binding sites across the *Firre* locus. CTCF has previously been shown to play a critical role in X-chromosome pairing and counting[48]. Similarly, this array of CTCF binding sites across the *Firre* locus might further facilitate inter-chromosomal interactions with the 5Mb X-chromosome localization domain. Finally, our study demonstrates that the formation of these cross-chromosomal interactions is altered upon genetic depletion of the *Firre* locus.

This model has several implications for potential new roles for lncRNA-mediated gene-regulation. For example, lncRNAs could bring genes involved in a similar biological process into close proximity allowing for co-regulation in space and time, serving as nuclear organization factors[19]. This appears to be the case for Firre, where several genes involved in energy metabolism and adipogenesis are organized together in spatial-proximity and are typically co-expressed, consistent with the previously described role of Firre in adipogenesis[28]. Underscoring the physiological relevance of such a model, either genetic deletion or transcriptional depletion of Firre results in the perturbation of cell physiology in both mESCs and adipocytes, respectively[28]. Future studies will require genetic studies in mouse models to further illuminate the role of Firre in mammalian development and disease.

Online Methods

Repetitive sequence analysis (FSA)

The sequence for Firre was scanned for repetitive elements using the *ab initio* repeat detection algorithm RepeatScout[47]. This sequence was then aligned back to the genome (mm9 or hg19) using BLAT[49] with the following parameters: "-stepSize=5 -repMatch=2253 -minScore=50 –minIdentity=0". Genomic DNA for the hits was extracted and multiply aligned with the Fast Statistical Aligner, a probabilistic multiple alignment tool specifically engineered to accommodate multiple alignment of sequences with potentially non-uniform evolutionary constraint[50]. Fast Statistical Alignment uses pair hidden Markov models to estimate gap and substitution parameters for the multiple alignment scoring function, improving alignment robustness.

Cloning *Firre*

Total RNA (1 μg) was reverse transcribed following the instructions in the Superscript III kit (Life Technologies, #18080-051). The thermocycling conditions were: 25°C for 10 minutes, 55°C for 1 hour, 70°C for 15 minutes and 4°C final. 2 μL of the cDNA was mixed with 21 μL of water, 2 μL of 10 μM primers and 25 uL 2x Phusion Mastermix (New England Biolabs, # M0531S). The PCR conditions were: 1) 98°C for 30 seconds, 2) 98°C for 10 seconds, 3) 66°C for 30 seconds, 4) 72°C for 3 minutes, 5) 72°C for 5 minutes, 6) and 4°C final, with 45 cycles repeating steps 2–4. The extension time varied with the length of the lincRNA. The products were checked on 1% agarose gel. Nested PCR was performed when necessary using purified PCR products instead of cDNA. Longer isoforms were gel purified and then subjected to the following cleaning steps. The PCR products were purified using SPRI beads (Beckman Coulter, #A63880), following the instructions in the manual. SPRI beads were added to the PCR product and incubated at room temperature for 2 minutes. The mix was put on a magnet for 4 minutes and the supernatant was removed. The beads were washed with 100 μL of 70% EtOH for 30 seconds twice and placed at 37°C for 5 minutes until the beads appeared dry. The PCR product immobilized on the beads was eluted with 30 μL of water on the magnet for 5 minutes.

The purified PCR product was quantified and used in BP reactions. The amount of DNA to be added was calculated as described in the Gateway cloning manual (Invitrogen). The BP reaction was set up according to the BP Clonase II instructions (Life Technologies, #11789020).

For transformations, 1 vial of Omnimax 2T1R (Life Technologies, #8540-03) cells were used for four BP reactions. The steps outlined in the Omnimax 2T1R manual were followed. The transformation plates were incubated at 37°C overnight, and the colonies were sequenced through Genewiz. When the inserts were verified, the plasmids were prepared using the Qiagen mini-prep kit (Qiagen, #27104).

In situ hybridization

The Firre probe was generated by PCR from adult brain cDNA and subcloned in pCRII-TOPO (Life Technologies, #K4610-20). The antisense riboprobe was generated by *in vitro* transcription using SP6 polymerase (Roche Applied Science, #10810274001) as previously described[51]. For non-radioactive *in situ* hybridizations, staged embryos were dissected in 1X PBS (Invitrogen) and fixed in 4% paraformaldehyde overnight at 4°C. For E14.5 cross-sections, embryos were washed overnight at 4°C in 30% sucrose/PBS followed by 1:1 ratio of 30% sucrose/OCT Clear Frozen Section Compound (VWR, #95057-838) for 1 hour. Embryos were then placed in fresh OCT, frozen and stored at −80°C until sectioning. Frozen serial sections, 20 μm thick, were prepared using a HM550 cryostat (Thermo Scientific) and

mounted onto Superfrost Plus slides (VWR #48311-703). Sections were permeabilized with 10 μg/ml proteinase K (Roche Diagnostics) for 10 minutes, washed with 1X PBS, treated 10 minutes in RIPA buffer and cross-linked again for 5 minutes in cold 4% paraformaldehyde. Sections were then pre-hybridized for 1 hour at room temperature at 70°C (50% formamide, 5X SSC, 5X Denhardts, 500 μg/ml Salmon sperm DNA, 250 μg/ml Yeast RNA) and then incubated overnight at 70°C in the same solution containing 2 μg/ml of DIG-labeled riboprobe. Sections were washed, blocked 1 hour with 10% sheep serum and incubated overnight at 4°C with 0.375 U/ml alkaline phosphatase-labeled anti-DIG antibody (Roche Diagnostics). Signal was detected by exposing sections to NBT-BCIP (Sigma #B1911), 0.1% Tween-20. Reaction was stopped with washes in 1X PBS supplemented with 0.1% Tween-20, and sections were mounted in Fluoromount-G (Southern Biotech #0100-01). The E14.5 whole embryo cross-section image (Fig. 1C) was taken from the Eurexpress Database (Assay ID #euxassay_013928, www.eurexpress.org). Non-radioactive *in situ* hybridizations of E15.5 embryo brains were performed on 40 μm vibratome sections (Leica) mounted on Superfrost Plus slides (VWR) using reported methods[52]. Sense probes were used as negative controls in all experiments. The PCR primers used to generate the Firre probe and the Firre riboprobe sequence used for the hybridizations are listed in Suppl Table 4:

Cellular fractionation

The cells grown in 15 cm dishes were washed with 5 mL of 1X PBS and trypsinized with 3 mL of TrypLE (Invitrogen) at 37°C for 3–5 minutes. The trypsin was quenched with 5 volumes of ice-cold growth media (DMEM (Invitrogen), 10% FBS (Invitrogen), 1% Pen-Strep (Invitrogen), and 1% L-Glutamine (Invitrogen)) and the cells were pelleted at 200xg for 3 minutes and resuspended in 1 ml of ice-cold 1X PBS. The resuspension was centrifuged at 200xg for 10 minutes at 4°C. The supernatant was carefully removed without disturbing the pellet; the remaining packed pellet volume was estimated for the next steps. The pellet was resuspended in 5 packed pellet volumes of ice-cold cytoplasmic extraction buffer (20 mM Tris, pH 7.6 (Ambion), 0.1 mM EDTA (Ambion), 2 mM MgCl2 (Ambion), 1X protease inhibitors (VWR), 0.5 U/μL RNaseOUT (Invitrogen)). The cells were incubated first at room temperature for 2 minutes then on ice for 10 minutes. The cells were lysed by adding CHAPS to a final concentration of 0.6%. The sample was then homogenized by passing it through a 1 ml syringe and centrifuged at 500x for 5 minutes at 4°C. The 70–80% of the supernatant was taken and saved at −80°C; this was the cytoplasmic fraction. The remaining supernatant was carefully removed and the pellet was washed with cytoplasmic extraction buffer supplemented with 0.6% (w/v) CHAPS. The sample was centrifuged at 500xg for 5 minutes at 4°C, and the entire supernatant was discarded. The wash step was repeated one more time. The pellet was then resuspended in 2 packed pellet volumes of nuclei suspension buffer (10 mM Tris, pH 7.5, 150 mM NaCl, 0.15% (v/v) NP–40, 1X protease inhibitors, 0.5 U/μL RNaseOUT). The nuclear suspension was layered on 5 packed pellet volumes of sucrose cushion (10 mM Tris, pH 7.5, 150 mM NaCl, 24% (w/v) Sucrose, 1X protease inhibitors, 0.5 U/μL RNaseOUT) and pelleted at 14,000 rpm for 10 minutes at 4°C. The supernatant was discarded and the pellet was washed with 10 packed pellet volumes of ice-cold 1X PBS supplemented with 1 mM EDTA. The sample was then centrifuged at 500xg for 5 minutes at 4°C. The pellet constituted the nuclear fraction.

Fluorescence *In Situ* Hybridization (FISH)

The FISH protocol was followed as described previously[29]. Briefly, oligonucleotide probes targeting and tiling the intron of Firre were conjugated to Alexa594 fluorophores, and the probes targeting and tiling the exon were conjugated to tetramethylrhodamine (TMR) and HPLC purified. Before the hybridization, the adherent cells were fixed (10 minutes with 4% formaldehyde) and permeabilized with 70% EtOH in two-chamber coverglasses. mESCs and hESCs were fixed in solution after they were collected from the plate: the cells were

incubated at room temperature in 2% formaldehyde solution for 10 minutes, followed by centrifuging at 1000xg for 3 minutes. The cells were washed with 1X PBS twice with centrifuging at 1000xg for 3 minutes in between. The cells were permeabilized with 70% EtOH. The ESCs were then plated on gelatinized coverglasses. Prior to the hybridization, the cells were rehydrated with the wash buffer containing 10% formamide (Ambion, #AM9342) and 2X SSC (Ambion, #AM9765) for 5 minutes. Then the probes (0.5 ng/μL final) were hybridized in 10% dextran sulfate (Sigma, #D8906), 10% formamide, and 2x SSC at 37°C overnight. After hybridization, the cells were washed in wash buffer at 37°C for 30 minutes twice (with the addition of DAPI in the second wash), followed by washing with 2X SSC twice. The imaging was done immediately after using 2X SSC as the mounting medium.

The same protocol was followed for the co-FISH experiments. The probes targeting and tiling the introns of the trans sites (Slc25a12, Ypel4, Ppp1r10) were conjugated to Quasar570. Co-FISH assays were conducted as indicated in wild type male mESC, *ΔFirre* male mESC, or mLF. Quasar670 was used as an additional fluorophore when working with three colors and trans sites: Quasar570 and Quasar670 were used together when staining for trans sites.

Actinomycin-D treatment

Actinomycin-D (Act-D) (Sigma, # A9415-2MG) was resuspended in DMSO with a final concentration of 2 mg/mL. Act-D was thoroughly mixed with the 2I media (2 μg/ml) and added on the male mESCs at 0, 1.5, 3, and 6 hours.

RNA Antisense Purification (RAP) Analysis

RAP was performed as described[19]. Briefly, the RNA of interest is tiled with 120 bp antisense nucleotides that have been biotinylated. Two distinct pools of antisense probes targeting Firre and one pool containing sense probes (negative control) were generated. The hybridization was done in duplicate crosslinked and precleared lysates with 20 ng (350 fmol) of oligos. The oligos were then captured by streptavidin beads and, the elutions for RNA and DNA were performed. Consistent with standard ChIP-Seq assays duplicate pull-downs were performed and sequenced to control for technical variability.

For X-chromosome enrichment analysis (Fig. 3A), the X-chromosome was divided into 10Kb bins and a linear regression of counts per bin was performed against each replicate and the input control. The slope of the linear regression was used as a normalization factor (alpha) between the two libraries. Enrichment levels relative to input were calculated by dividing the experimental counts for each bin by the input counts times alpha.

To identify significant regions bound by the Firre RNA *in trans*, we used the Scripture peak-calling algorithm[53] to call significant peaks across each of the replicate sequencing bam files, including the input control and anti-sense control. All peaks were merged using the Bedtools mergeBed[54] program to obtain the universe of significant peaks across all samples. A .gtf file of significant peaks, along with the replicate .bam files for each of the samples was used as input for Cuffdiff2[26] for quantification and differential testing. Cuffdiff2 was run using default parameters with the addition of the '--no-length-correction' argument to disable length correction. Significant peaks were called using the Cuffdiff2 test-statistic with $p<0.1$.

Retroviral Overexpression of Firre

The overexpression vector for overexpressing Firre was made by modifying the pLenti6.3/TO/V5-DEST (Snap Gene) destination vector. We modified by removing the WPRE, the

SV40 promoter and the blastacidin resistance gene, keeping the gateway tails the same, to prevent any interference with the lincRNA structure and function. All the transductions were done as follows: the cells were split into 12-well dishes and resuspended in media with 4 μg/mL polybrene. Immediately after, 100 μL of virus (of the same titer; if not the same titer, the volume was adjusted accordingly) was added to each well. The untransduced control was used to measure the overexpression levels by qRT-PCR. The sequences of the isoforms 1.1 (+RRD), 1.4 (+RRD), 1.6 (-RRD) used for mouse transductions are in Supplementary Table 1 (bed file) and the sequence of the human isoform used for HEK293 and HeLa transductions is given in Suppl. Table 4:

Cell Culture

HEK293 (ATCC: CRL-1573), HeLa (ATCC: CCL-2), mLF (ATCC: CCL-206) and hLF (ATCC: IMR90) cells were grown in growth media (see above) at 37°C at 5% CO_2. Male (Novus: NBP1-41162) and female (RIKEN: AES0010) mouse ES cells were grown in previously gelatinized (0.2%) dishes with 2I media containing 125 mL DMEM/F12 (Invitrogen), 83.5 μL BSA fraction V (50 μg/ml relative to DMEM) (Invitrogen, 15260-037, 75mg/ml), 125 mL Neurobasal medium (Invitrogen, 21103-049), 625 μL of the Ndiff Neuro2 (200X, relative to Neurobasal medium) (Millipore, SCM012), 2.5 mL B27 minus vitamin A (50X, relative to Neurobasal medium) (Invitrogen, 12587-010), 2 μL beta-mercaptoethanol, 1 μM PD0325901 (Stemgent, 04-0006), 3 μM CHIR99021 (Stemgent 04-0004), 25 μL LIF ESGRO (from Chemicon, ESG1106), 1% pen-strep (Invitrogen, 15140-163), 1% Non-Essential Amino Acids (Invitrogen, 11140-076), 1% L-glutamine (Invitrogen, 25030-164). The plating density of mES cells was chosen to be 30,000–50,000/cm^2. Adipocytes were grown as described previously [28].

Pull-down with *in vitro* biotinylated RNA

The cloned in transcript in pdest14 plasmid vector was linearized with Nhe1. Phenol-chloroform extracted and ethanol precipitated template was then used in *in vitro* transcription, which included: 20 ug/ml DNA template, 40 mM Tris (pH 7.9), 2.5 mM Spermidine, 26 mM $MgCl_2$, 0.01% Triton X-100, 8 mM GTP, 5 mM ATP, 5 mM CTP, 1.3 mM UTP, 0.7 mM Bio-16-UTP (Epicentre), 5 mM DTT, 20 mM $MgCl_2$, RNAseOut 80 U/ml, 20 U of T7 RNA polymerase (Life Technologies, #18033-019). The mix was incubated at 37°C until a white precipitate formed. After the reaction reached completion, 60 mM-final EDTA was added to dissolve the precipitate. Due to the biotin, RNA will partition into the organic layer if it is phenol-extracted. Therefore, the *in vitro* transcriptions were first treated with DNase (Worthington, #LS006353) (37°C for 10 minutes followed by EDTA addition and 75°C for 10 minutes), then cleaned with the Bio-Spin 30 columns (BIO-RAD, # 732-6231).

For end labeling, the *in vitro* transcribed RNA (without labeled UTPs) was treated with the 3′ end biotinylation kit of Thermo Scientific Pierce (#20160).

The lysate was prepared by lysing 15 cm dishes into 1 ml of lysis buffer (150 mM KCl, 25 mM TRIS- HCl pH 7.4, 5mM EDTA, 5 mM $MgCl_2$, 1% NP-40, 1X protease inhibitor, 0.5 mM DTT, 100 U/ml RNAseOut) for 30 minutes at 4°C. The lysate was centrifuged at 13,000 rpm for 30 minutes and filtered with 0.45 um filter. The concentration of the lysate was measured by BCA protein assay (Thermo Scientific, # 23225).

For the pull-down, 1.5 mg of the lysate was initially pre-cleared with the Magnetic MyOne Streptavidin T1 beads (Life Technologies, #65601) for 30 minutes at 4°C. The beads were prepared as described in the manual. The precleared lysate was 2X diluted and supplied with 0.1 μg/μl tRNA, to which 30 pmoles of biotinylated RNA was added. The RNA was

incubated in the lysate for 2 hours at 4°C rocking, after which 40 μl of MyOne Streptavidin T1 beads was added to the mix. The mix was incubated for another hour at 4°C. The beads were washed for 3 times (10 minutes each) with 1 ml of the wash buffer (lysis buffer but with 300 mM KCl) on a magnetic rack. Finally, the beads were resuspended in 30 μl of sample buffer (4X, Biorad) and reducing agent (20X, Biorad) and boiled for 5 minutes at 95°C. The samples were then run on 4–12% gradient Bis-Tris gels and stained with Sypro Red as described (Life Technologies, #S-12000) for protein detection and mass spectrometry or were transferred to a PVDF membrane for Western blotting.

The mass spectrometry analysis was done as follows: The eluates from RNA pull-downs done in three different cellular contexts (mouse adipose tissue, mouse adipocyte, and mESC lysates) using five different RRD+ isoforms and one RRD– isoform were run on a gel as described. The bands that are differential between RRD+ and RRD– isoforms were cut and processed for mass spectrometry. To identify proteins that preferentially co-precipitated with Firre in an RRD-dependent manner, we took the difference between the peptide counts of the RRD+ and RRD– isoforms. We repeated the differential analysis for each of the RRD+ isoforms and took the top 10% of the differential peptide count scores for each isoform identified in both mESC and adipocyte lysates.

The transfer for Western blotting was done in transfer buffer that was prepared in the following ratios: 100 ml of 10X TG (Biorad), 200 ml methanol and 700 ml ddH_2O. The membrane was activated in methanol first and equilibrated in transfer buffer prior to transfer. The transfer was done at 70 watts for 1 hour. After the transfer, the membrane was washed with methanol for blocking and incubated with the primary antibody diluted in 1:1000 in 0.1% Tween, 1% non-fat milk, and 1X PBS for 4–5 hours at RT or 4°C O/N. Following the primary antibody incubation (human hnRNPU (3G6): SantaCruz sc-32315 (reactivity with the species is validated and shown on the Santa Cruz website), mouse hnRNPU: Abcam ab20666 (validated by [55]), the membrane was washed 3 times with wash buffer (1X PBS supplemented with 0.1% Tween) and then incubated with the secondary antibody for 1 hour at RT. The membrane was washed again 3 times with wash buffer and then developed using SuperSignal West Chemiluminescent Thermo Scientific reagents. 20% of the input lysate was used for all the Western blots (Fig. 6A, B, Supplementary Fig. 5D).

Endogenous RNA pull-down

The 23–25 bp oligos (sequences in Suppl. Table 4) were synthesized with 18S linker and desthiobiotin at the 5′ end through IDT. The same protocol for RNA-pull downs above was followed, except instead of adding the biotinylated RNA, the DNA oligos were added to the lysate and an annealing step was followed. Annealing was done by incubating the lysate and the oligos at 37°C for 15 minutes, at room temperature for 15 minutes, and at 4°C for 6 hours-O/N. The rest of the steps were the same as above except for these: 1) during the incubation with the beads, heparin was spiked in at 0.5 μg/μL in the last half hour, 2) in the first wash 0.5 μg/μL heparin was spiked in again, and 3) instead of boiling the beads, the RNA and protein were eluted from the beads with 12.5 mM Biotin for 30 minutes at RT and 3 hours at 4°C. The 60% of the elution was used to extract RNA and 40% to run a protein gel.

RNA Immunoprecipitation

The protein lysate (1.5 mg) was incubated with 6–8 μg of the hnRNPU or IgG antibody (hnRNPU (3G6): Santa Cruz sc-32315, hnRNPU (H-94): Santa Cruz sc-25374 (validation shown on the Santa Cruz website), Anti-hnRNPU: Abcam ab20666, Mouse (G3A1) IgG1: Cell Signaling 5415) at 4°C for 2–3 hours. Then 45 μL of protein G Dynabeads (Life Technologies, #10003D) that were previously washed twice in 500 μL lysis buffer were

added to the lysate and antibody mix. The lysate, antibody, and beads were incubated at 4°C for another 2 hours. The beads were washed 3 times (10 minutes each) with 1 mL of lysis buffer. The RNA was extracted by adding 1 mL of TRIzol (Life Technologies, #15596-018) to the beads. For the total input RNA, 10% of the input lysate was mixed with 1 mL of TRIzol. For 1 ml of TRIzol, 200 μL of chloroform was added, and the mix was centrifuged at 4°C at 13,000 rpm for 15 minutes. The aqueous layer was then added 1 volume of isopropanol, 1/10 volume KOAc, and 1 μL of glycoblue and kept at −20°C for at least one hour. The samples were then centrifuged at 13,000 rpm at 4°C for 30 minutes. The supernatant was removed and the pellet was washed with 1 mL of ice-cold 70% EtOH twice (centrifuging 2 minutes each time at 4°C). The pellet was then resuspended in 15 μL of RNase-free water.

CLIP-Seq Analysis

We analyzed two human hnRNPU CLIP-Seq datasets generated independently by Huelga et al. (GSE34993)[42] and Xiao et al. (GSE34491)[43]. We downloaded fastq files and aligned using TopHat[56] to hg19 and a custom transcriptome GTF consisting of UCSC coding genes, a recently published lncRNA catalog[57], and cloned Firre isoforms. We used the RNA-Seq differential expression software Cuffdiff to estimate read counts to Firre in all CLIP and total RNA datasets[26]. We performed a Poisson-based statistical test for enrichment of aligned reads in the CLIP versus total RNA.

RNAi-mediated Knockdown of hnRNPU

mESCs were transfected by the reverse transfection method in 6-well plate format. Briefly, the lipofectamine RNAiMAX (6 μl/well, Life Technologies. # 13778030) and siRNA (50 nM final, Dharmacon On-Targetplus smart pool for mouse (L-051574-01-0005) and for human (L-013501-00-0005)) complexes were prepared in 400 μl of Opti-MEM and incubated at room temperature for 20–30 minutes, during which the mESCs were prepared for splitting. The split was done as follows: the cells were washed with 1X PBS and trypsinized for 3 minutes. The trypsin was quenched by 2I media; the cells were centrifuged for 5 minutes at 850 rpm at 4°C, resuspended in new 2I media, and counted. Approximately 280,000 cells were plated for each well of a 6-well plate. Immediately after, the Lipofectamine/oligo complexes were added to the wells. The media was changed after 24 hours and the cells were harvested after 96 hours for the complete knockdown of the protein (checked by qRT-PCR and by Western blot). The same protocol was used for HEK293s and HeLa cells; however, they were plated at a density of ~21,000 cells/cm^2.

RNA Extraction

RNA extraction was performed by adding 1 ml of TRIzol to each well of a 6-well plate. 200 μL of chloroform was added, and the mix was centrifuged at 4°C at 13,000 rpm for 15 minutes. The aqueous layer was processed on the RNeasy Mini columns (Qiagen, #74104). The RNA was reverse transcribed using the SuperScriptIII First-Strand Synthesis kit. The cDNA synthesis was performed at 25°C for 5 minutes, 50°C for 1 hour, and 70°C for 15 minutes. Then the cDNA (15 ng per well of the 384 qPCR plate) was added 1:1 to the SYBR and primer mix (100 nM) for qRT-PCR. All the primers used in qRT-PCR are shown in Supplementary Table 4.

RNA-Seq library preparation, sequencing, and analysis

200ng of extracted RNA from each of 2 Δ*Firre* and 3 wild-type JM8A male mouse ES cell cultures was used as input for the Illumina TruSeq library preparation kit, using manufacturer's guidelines. Libraries were individually barcoded and library size distribution and quality were assayed using a DNA High-sensitivity Chip on the Agilent Bioanaylzer

2100. Libraries were pooled and paired-end 35bp fragments were generated on an Illumina MiSeq sequencer to an average depth of 9 Million fragments per sample. Fragments were aligned to the mouse genome (mm9) using Tophat2[58] with default options and the UCSC transcriptome as a reference. Aligned reads were quantified against all mouse UCSC genes using Cuffdiff2[59] with default options. Significantly differentially expressed (DE) genes were selected with an FDR of 5%. Circular representation of DE gene projections onto mm9 (Figure 5f) was generated using the Circos utility (http://circos.ca/). Since RNA-Seq data are heteroscedastic not normally distributed, we chose to evaluate pathway enrichment by using a preranked GSEA analysis. This was conducted by using the GSEA tool[60] with a list of all genes ranked by cuffdiff2 test statistic (KO/WT), against the c2.cp.reactome.v4.0.symbols.gmt gene set collection (MSigDB, Broad). Gene sets were selected as significantly enriched if the nominal p-value was less than 0.01.

Targeting and generation of conditional *ΔFirre* mESC

To generate ESCs specifically deficient in Firre, a two-step targeting strategy was used to introduce loxP sites in the 5′ and 3′ ends of the *Firre* locus. Targeting of only one allele was needed to obtain *Firre*-deficient ESCs as *Firre* resides on the X-chromosome and male ESCs (JM8) were used. To generate the *Firre* 3′ targeting construct we amplified the homology arms (primer sequences in Suppl Table 4), which were then cloned into the pEASY-FLIRT vector:

In this targeting construct, the neo cassette that serves as a selection marker during the targeting process was flanked by two Frt sites. To generate the *Firre* 5′ targeting construct, we amplified the homology arms (primers in Suppl Table 4), which were then cloned into the newly generated pEASY-Hygro vector:

In this construct, the hygro cassette that serves as a selection marker during the targeting process was flanked by two Loxp sites.

ES cells were maintained under standard conditions and targeted as previously described[30]. In brief, the vector targeting the 3′ end of *Firre* was electroporated into C57BL/6 ES cells (JM8) and grown under selection with neomycin. Homologous recombinant ES cells identified by PCR analysis were subsequently electroporated with the vector targeting the 5′ end of *Firre* and grown under selection with hygromycin. Double-targeted ES cells were identified by PCR analysis. To delete *Firre*, double-targeted ES cells were electroporated with a Cre recombinase expressing plasmid (pGK-Cre-bPA). PCR genotyping was used to identify clones, in which *Firre* had been deleted.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Biosearch for materials, advice and help with FISH experiments, especially M. Beal, H. Johansson, A. Orjalo, S. Coassin and R. Cook. We are grateful to the entire Rinn and Raj labs for the generous help with experiments, bioinformatics, and preparation of the manuscript. This work was supported by 1DP2OD00667 (J.L.R.), P01GM099117 (J.L.R), 1DP20D008514 (A.R), and P50HG006193-01 (J.L.R.) and the Howard Hughes Medical Institute (R.A.F.).

References

1. Huarte M, Rinn JL. Large non-coding RNAs: missing links in cancer? Hum Mol Genet. 2010; 19:R152–61. [PubMed: 20729297]

2. Rinn JL, Chang HY. Genome Regulation by Long Noncoding RNAs. Annu Rev Biochem. 2012; 81:145–66. [PubMed: 22663078]
3. Carninci P, et al. The transcriptional landscape of the mammalian genome. Science. 2005; 309:1559–63. [PubMed: 16141072]
4. Derrien T, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 2012; 22:1775–89. [PubMed: 22955988]
5. Batista PJ, Chang HY. Cytotopic localization by long noncoding RNAs. Curr Opin Cell Biol. 2013; 25:195–9. [PubMed: 23279909]
6. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. Nature. 2012; 482:339–46. [PubMed: 22337053]
7. Clemson CM, et al. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol Cell. 2009; 33:717–26. [PubMed: 19217333]
8. Maass PG, et al. A misplaced lncRNA causes brachydactyly in humans. J Clin Invest. 2012; 122:3990–4002. [PubMed: 23093776]
9. Nie L, et al. Long non-coding RNAs: versatile master regulators of gene expression and crucial players in cancer. Am J Transl Res. 2012; 4:127–50. [PubMed: 22611467]
10. Pandey RR, et al. Kcnq1ot1 Antisense Noncoding RNA Mediates Lineage-Specific Transcriptional Silencing through Chromatin-Level Regulation. Molecular Cell. 2008; 32:232–246. [PubMed: 18951091]
11. Rinn JL, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell. 2007; 129:1311–23. [PubMed: 17604720]
12. Tsai MC, et al. Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. Science. 2010; 329:689–693. [PubMed: 20616235]
13. Vallot C, et al. XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells. Nat Genet. 2013
14. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science. 2008; 322:750–6. [PubMed: 18974356]
15. Kaneko S, Son J, Shen SS, Reinberg D, Bonasio R. PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. Nat Struct Mol Biol. 2013; 20:1258–64. [PubMed: 24141703]
16. Davidovich C, Zheng L, Goodrich KJ, Cech TR. Promiscuous RNA binding by Polycomb repressive complex 2. Nat Struct Mol Biol. 2013; 20:1250–7. [PubMed: 24077223]
17. Jeon Y, Sarma K, Lee JT. New and Xisting regulatory mechanisms of X chromosome inactivation. Current Opinion in Genetics & Development. 2012; 22:62–71. [PubMed: 22424802]
18. Plath K, Mlynarczyk-Evans S, Nusinow DA, Panning B. Xist RNA and the mechanism of X chromosome inactivation. Annu Rev Genet. 2002; 36:233–78. [PubMed: 12429693]
19. Engreitz JM, et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science. 2013; 341:1237973. [PubMed: 23828888]
20. Simon MD, et al. High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. Nature. 2013; 504:465–9. [PubMed: 24162848]
21. Bouvier D, Hubert J, Seve AP, Bouteille M. Nuclear Rna-Associated Proteins and Their Relationship to the Nuclear Matrix and Related Structures in Hela-Cells. Canadian Journal of Biochemistry and Cell Biology. 1985; 63:631–643. [PubMed: 4041964]
22. Nickerson JA, Krochmalnic G, Wan KM, Penman S. Chromatin architecture and nuclear RNA. Proc Natl Acad Sci U S A. 1989; 86:177–81. [PubMed: 2911567]
23. Pederson T, Bhorjee JS. Evidence for a role of RNA in eukaryotic chromosome structure. Metabolically stable, small nuclear RNA species are covalently linked to chromosomal DNA in HeLa cells. J Mol Biol. 1979; 128:451–80. [PubMed: 571474]
24. Umlauf D, Fraser P, Nagano T. The role of long non-coding RNAs in chromatin structure and gene regulation: variations on a theme. Biol Chem. 2008; 389:323–31. [PubMed: 18225988]
25. Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. Genes Dev. 2009; 23:1494–504. [PubMed: 19571179]

26. Mao YS, Sunwoo H, Zhang B, Spector DL. Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. Nat Cell Biol. 2011; 13:95–101. [PubMed: 21170033]
27. Delpretti S, et al. Multiple enhancers regulate Hoxd genes and the Hotdog LncRNA during cecum budding. Cell Rep. 2013; 5:137–50. [PubMed: 24075990]
28. Sun L, et al. Long noncoding RNAs regulate adipogenesis. Proc Natl Acad Sci U S A. 2013
29. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA molecules using multiple singly labeled probes. Nature Methods. 2008; 5:877–879. [PubMed: 18806792]
30. Cuddapah S, et al. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Res. 2009; 19:24–32. [PubMed: 19056695]
31. Chu C, Quinn J, Chang HY. Chromatin isolation by RNA purification (ChIRP). J Vis Exp. 2012
32. Simon MD, et al. The genomic binding sites of a noncoding RNA. Proc Natl Acad Sci U S A. 2011; 108:20497–502. [PubMed: 22143764]
33. Mariner PD, et al. Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. Molecular Cell. 2008; 29:499–509. [PubMed: 18313387]
34. Lee EK, et al. miR-130 Suppresses Adipogenesis by Inhibiting Peroxisome Proliferator-Activated Receptor gamma Expression. Molecular and Cellular Biology. 2011; 31:626–638. [PubMed: 21135128]
35. Nukitrangsan N, et al. Effect of Peucedanum japonicum Thunb on the Expression of Obesity-Related Genes in Mice on a High-Fat Diet. Journal of Oleo Science. 2011; 60:527–536. [PubMed: 21937852]
36. Rubi B, del Arco A, Bartley C, Satrustegui J, Maechler P. The malate-aspartate NADH shuttle member Aralar1 determines glucose metabolic fate, mitochondrial activity, and insulin secretion in beta cells. Journal of Biological Chemistry. 2004; 279:55659–55666. [PubMed: 15494407]
37. Seo J, et al. Atf4 Regulates Obesity, Glucose Homeostasis, and Energy Expenditure. Diabetes. 2009; 58:2565–2573. [PubMed: 19690063]
38. Choy L, Derynck R. Transforming growth factor-beta inhibits adipocyte differentiation by Smad3 interacting with CCAAT/enhancer-binding protein (C/EBP) and repressing C/EBP transactivation function. Journal of Biological Chemistry. 2003; 278:9609–9619. [PubMed: 12524424]
39. Gohring F, Fackelmayer FO. The scaffold/matrix attachment region binding protein hnRNP-U (SAF-A) is directly bound to chromosomal DNA in vivo: A chemical cross-linking study. Biochemistry. 1997; 36:8276–8283. [PubMed: 9204873]
40. Hasegawa Y, et al. The Matrix Protein hnRNP U Is Required for Chromosomal Localization of Xist RNA. Developmental Cell. 2010; 19:469–476. [PubMed: 20833368]
41. Lobov IB, Tsutsui K, Mitchell AR, Podgornaya OI. Specificity of SAF-A and lamin B binding in vitro correlates with the satellite DNA bending state. Journal of Cellular Biochemistry. 2001; 83:218–229. [PubMed: 11573239]
42. Huelga SC, et al. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. Cell Rep. 2012; 1:167–78. [PubMed: 22574288]
43. Xiao R, et al. Nuclear matrix factor hnRNP U/SAF-A exerts a global control of alternative splicing by regulating U2 snRNP maturation. Mol Cell. 2012; 45:656–68. [PubMed: 22325991]
44. Wong LH, et al. Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. Genome Research. 2007; 17:1146–1160. [PubMed: 17623812]
45. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One. 2010; 5:e13984. [PubMed: 21085593]
46. Lee BK, Iyer VR. Genome-wide studies of CCCTC-binding factor (CTCF) and cohesin provide insight into chromatin structure and regulation. J Biol Chem. 2012; 287:30906–13. [PubMed: 22952237]
47. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. Bioinformatics. 2005; 21 (Suppl 1):i351–8. [PubMed: 15961478]

48. Donohoe ME, Silva SS, Pinter SF, Xu N, Lee JT. The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. Nature. 2009; 460:128–U147. [PubMed: 19536159]
49. Kent WJ. BLAT--the BLAST-like alignment tool. Genome Res. 2002; 12:656–64. [PubMed: 11932250]
50. Bradley RK, et al. Fast Statistical Alignment. Plos Computational Biology. 2009; 5
51. Arlotta P, et al. Neuronal subtype-specific genes that control corticospinal motor neuron development in vivo. Neuron. 2005; 45:207–221. [PubMed: 15664173]
52. Tiveron MC, Hirsch MR, Brunet JF. The expression pattern of the transcription factor Phox2 delineates synaptic pathways of the autonomic nervous system. J Neurosci. 1996; 16:7649–60. [PubMed: 8922421]
53. Guttman M, et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol. 2010; 28:503–10. [PubMed: 20436462]
54. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–2. [PubMed: 20110278]
55. Vizlin-Hodzic D, Johansson H, Ryme J, Simonsson T, Simonsson S. SAF-A Has a Role in Transcriptional Regulation of Oct4 in ES Cells Through Promoter Binding. Cellular Reprogramming. 2011; 13:13–27. [PubMed: 21235343]
56. Trapnell C, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012; 7:562–78. [PubMed: 22383036]
57. Kelley D, Rinn J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biol. 2012; 13:R107. [PubMed: 23181609]
58. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009; 25:1105–11. [PubMed: 19289445]
59. Trapnell C, et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 2013; 31:46–53. [PubMed: 23222703]
60. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102:15545–50. [PubMed: 16199517]

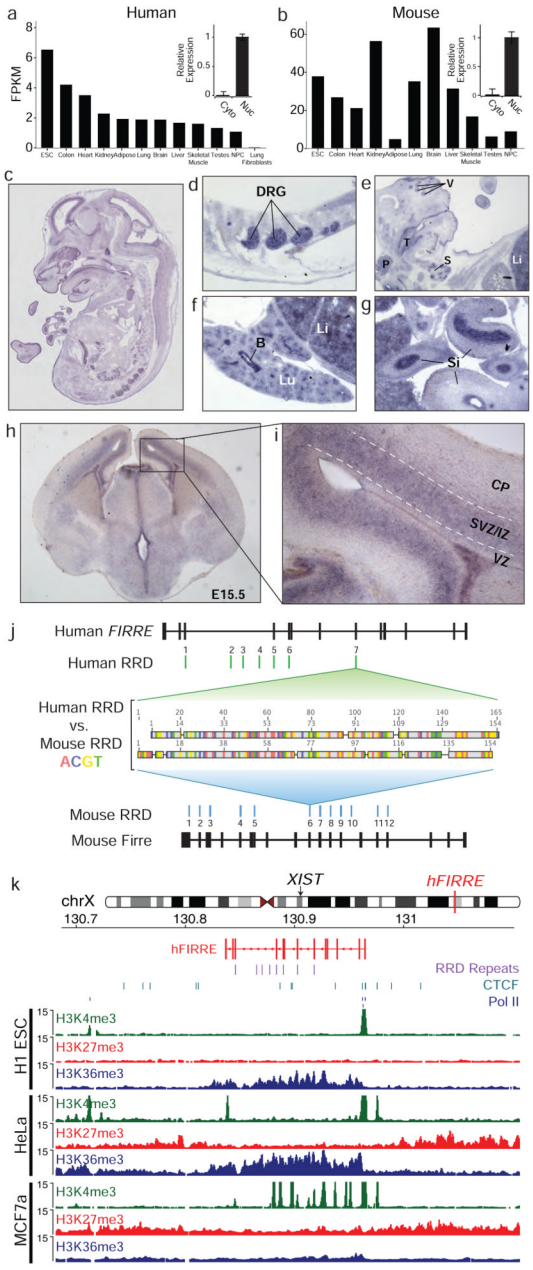

Figure 1. Firre is a novel, X-chromosome-localized lincRNA

(a, b) RNA sequencing abundance estimates in pair-matched mouse and human samples, including the cellular fractionations followed by qRT-PCR for Firre (insets). **(c–g)** Firre *in situ* hybridizations at E14.5 showing (**c**) the whole embryo, **(d)** dorsal root ganglia (DRG), (**e**) developing vibrissae (V), tongue (T), pituitary gland (P), salivary gland (S), (**f**) fetal liver (Li), lung (Lu), bronchi (B), and (**g**) small intestine (Si). **(h–i)** Firre *in situ* hybridizations at E15.5 in the proliferative ventricular (VZ) and subventricular zones (SVZ). **(j)** Transcript structure of human and mouse Firre with RRD marked in green and blue, respectively. **(k)** Top, the human Firre locus. Bottom, CTCF (light blue), H3K4me3 (green), H3K27me3 (red), and H3K36me3 (dark blue) tracks for the human *FIRRE* locus in human ES, HeLa and MCF7 cell lines. Chromatin modification tracks are expressed as raw counts.

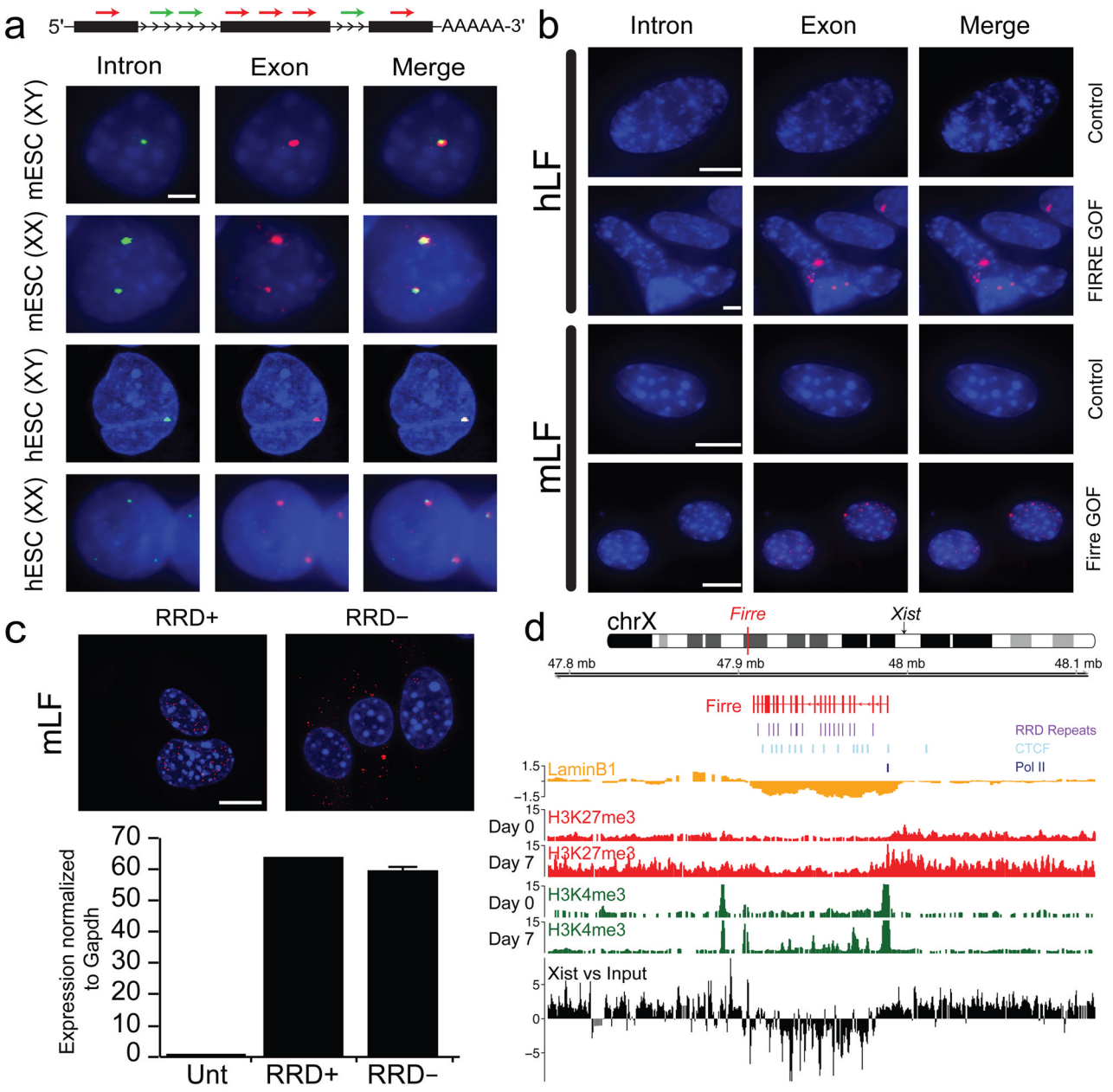

Figure 2. Firre is a novel, strictly-nuclear lincRNA that escapes X-chromosome inactivation

(a–c) Single molecule RNA FISH of introns in “green” (A594), exons in “red” (Cy3), and nuclei by DAPI. (**a**) Male and female human and mouse ESCs. Scale bar, 20 μm. **(b)** Viral overexpression of Firre in human and mouse lung fibroblasts (hLF, mLF) that do not express Firre. Scale bars, 15 μm. **(c)** Viral overexpression of Firre isoforms with or without RRD in mLFs. Scale bar, 15 μm. Bottom, quantification of expression; unt is untransfected control. Error bars, s.d., n=3. **(d)** The mouse *Firre* locus. The unique repetitive domain RRD (purple), CTCF (blue), LaminB1 (orange), H3K27me3 (red) and H3K4me3 (green), and Xist RAP (black). LaminB1 and Xist plotted as log fold-change on the y-axis relative to input; chromatin modifications as raw counts.

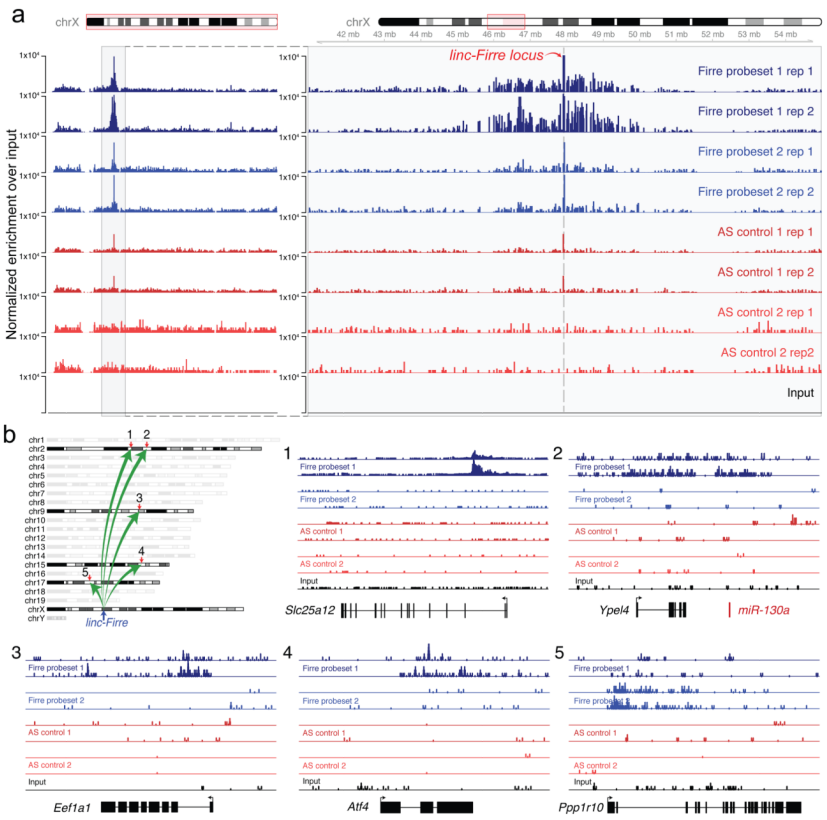

Figure 3. Firre forms cis and several trans chromosomal contacts

(a) RNA Antisense Purification (RAP) by Firre shown along the X-chromosome in male mESCs as fold enrichment relative to input after normalization for library depth. The genomic ~5Mb region flanking the Firre locus is highlighted. Replicate enrichment tracks normalized to the input (black) are presented for each of two Firre-targeting probes (blue tracks) and two sets of sense probes (negative controls) (red tracks). **(b)** RAP by Firre shown for 5 distinct inter-chromosomal genomic loci (1–5). Local genomic contexts are presented for each of the trans-interacting sites and their adjacent genes: *Slc25a12, Ypel4, Eef1a1, Atf4, and Ppp1r10*. Counts for the trans-chromosomal contacts of Firre are shown after normalization for sequencing depth.

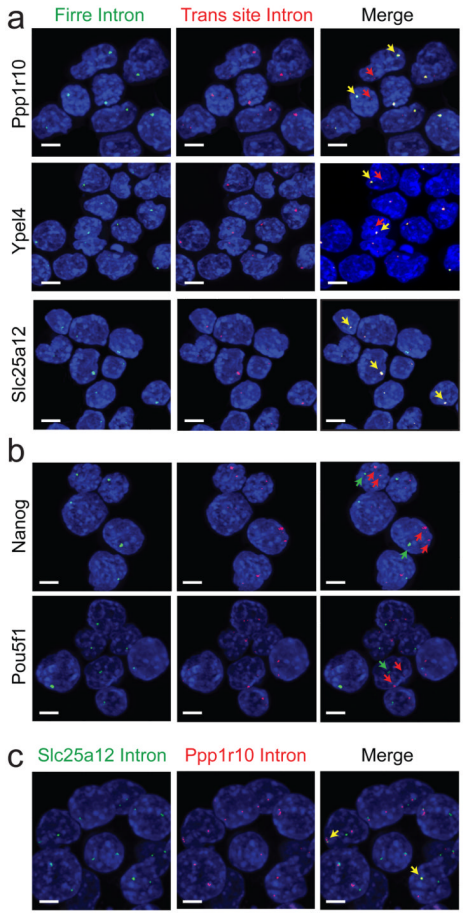

Figure 4. The trans-chromosomal contacts of Firre are co-localized within the nucleus

(a) Co-RNA-FISH: introns of Firre are labeled in green, the introns of its trans targets in red, and the nuclei by DAPI. Co-localization of Firre with Ppp1r10, Ypel4, and Slc25a12. **(b)** Co-RNA FISH of Nanog and Oct4. Green and red arrows indicate non-overlapping genomic loci, and yellow arrows indicate co-localization of both loci. **(c)** RNA FISH Co-localization of the trans-interacting loci Ppp1r10 and Ypel4. Scale bars, 40 μm.

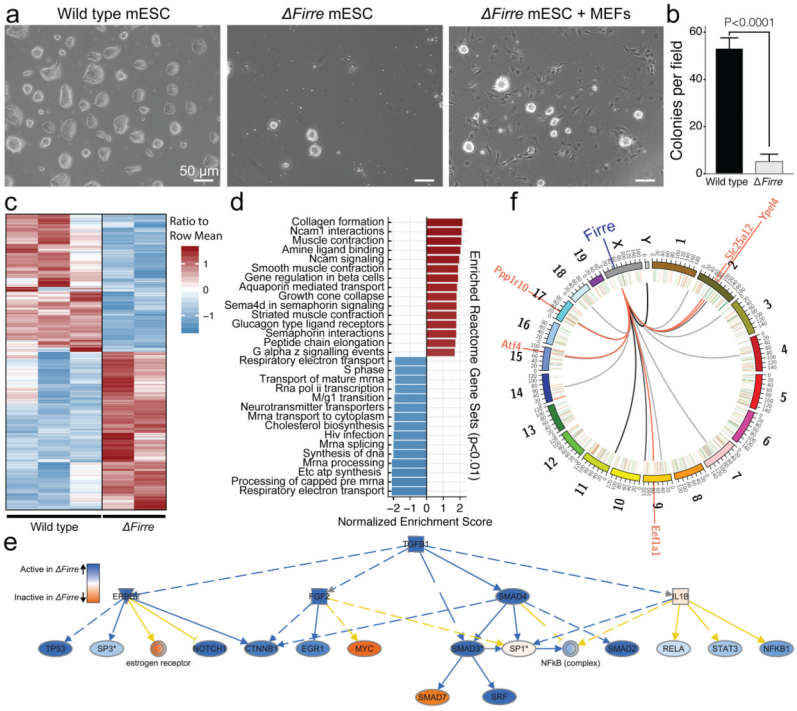

Figure 5. Characterization of Firre knockout in male mESCs

(a–b) Genetic deletion of Firre in male mESC (Δ*Firre)* with (+MEF) and without feeder cells. (**a**) images and (**b**) quantification of Δ*Firre* in mESCs. **(c)** Heatmap of 892 significantlly differentially expressed genes (Cuffdiff2; 1%FDR) between wild type (WT) and Δ*Firre* male mESCs. **(d)** Top 15 enriched and depleted significant (P<0.001; Mann-Whitney U-test). Reactome gene sets from a pre-ranked GSEA analysis on Cuffdiff2 test statistics. **(e)** Ingenuity Pathway Analysis mechanistic network diagram of significant ($p<6.31\times10^{-44}$) increase in predicted downstream Tgfβ signaling activity in the Δ*Firre* male mESC relative to WT. **(f)** Circos diagram of significant Firre RAP peaks (links) interacting with the *Firre* genomic locus (blue) in male mESCs. Peaks intersecting genic regions are highlighted in red and specifically labeled. Log_2 fold changes for significant (Cuffdiff2; 1% FDR) differentially expressed genes (Δ*Firre*/WT) are inscribed at corresponding genomic locations within the circle.

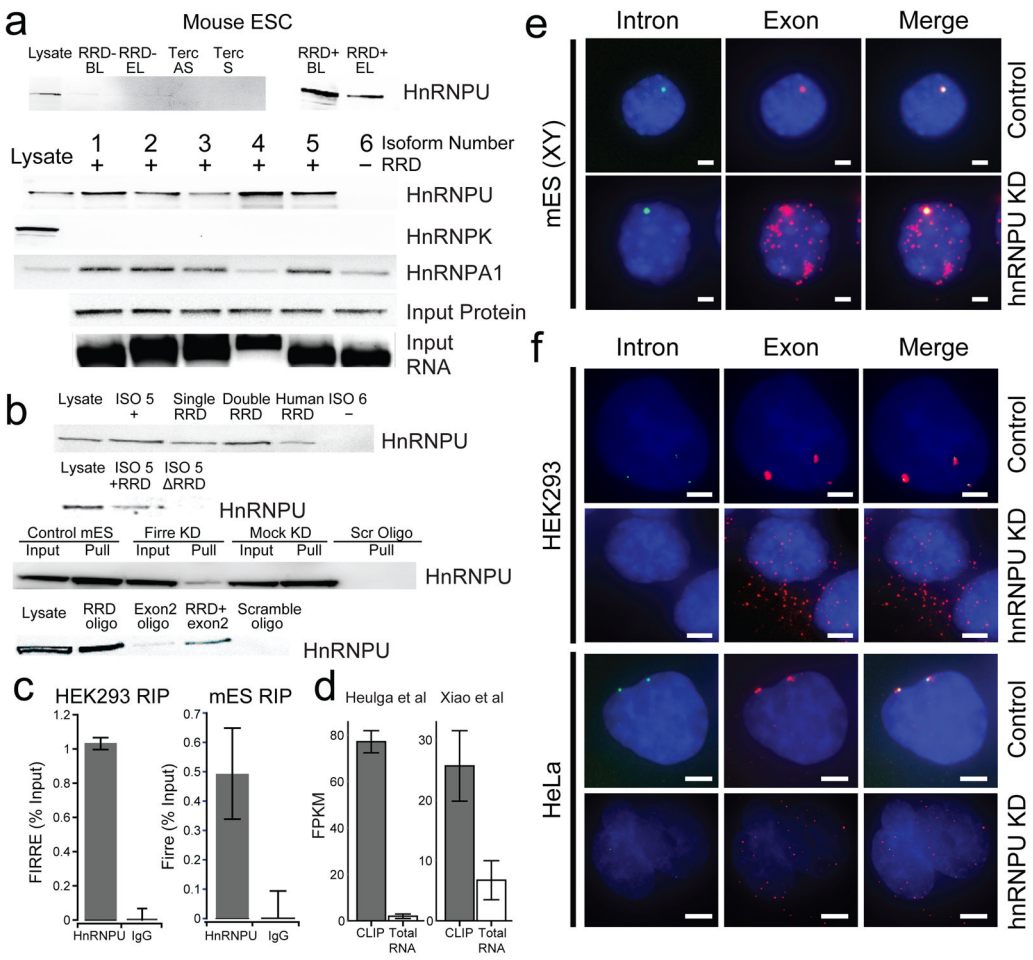

Figure 6. HnRNPU binds to the RRD of Firre and regulates its nuclear localization

(a) Western blots for RNA pull-downs: five with RRD and one without. Biotin end labeling (EL) versus body labeling (BL) for hnRNPU interaction is shown. TERC antisense (AS) and sense (S) are negative controls. hnRNPK and A1 are shown for RRD specificity. 20% of lysate was loaded. Input protein lysate and RNA shown as loading controls. **(b)** Western blots for RNA pull-downs performed with: synthetic RRD constructs (single, double, and human RRD) in mESCs (panel 1); RRD deleted isoform in mESCs (panel 2). Western blots for endogenous RNA pull-downs: desthiobiotin-DNA oligos complementary to Firre (top) compared to non-targeting scramble oligos in mESCs and HEK293s (bottom). **(c)** RIP with hnRNPU in HEK293s and mESCs shown as a percentage of input. Error bars are s.d., n=3. **(d)** CLIP data as gene level FPKM values. Error bars are s.d., n=3. **(e, f)** RNA-FISH targeting Firre in mESCs, HEK293s, and HeLa cells in the absence of hnRNPU: introns (green); exons (red); nuclei (blue). Scale bars, E: 20 μm; F: 5 μm. See Supplementary Fig. 7 for original images for a–b.

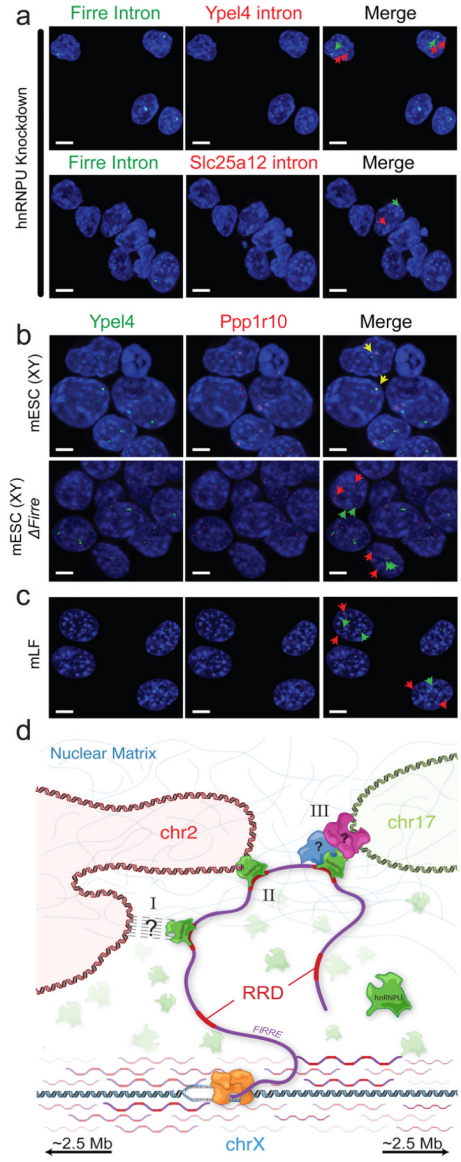

Figure 7. The trans-chromosomal contacts of Firre are mediated via its interaction with the nuclear matrix protein hnRNPU

(a) RNA-FISH co-localization of Firre introns (green) with Ypel4 or Slc25a12 introns (red) in the absence of hnRNPU. **(b)** RNA-FISH co-localization of the trans-interacting loci Ppp1r10 and Ypel4 in the absence of Firre expression in *ΔFirre* male mESCs compared to the wild-type mESCs (rows 1, 2). **(c)** RNA-FISH co-localization of trans-sites in mLFs (row 3). Scale bars, 40 μm. **(d)** A model for Firre as a 'regional organization factor.' Firre transcripts accumulate at the site of their transcription. hnRNPU binds to the RRD of Firre and facilitates interactions with trans-chromosomal regions through one of several possible mechanisms: I) Tertiary interactions with nuclear matrix components, II) Direct binding of hnRNPU to matrix attachment regions *in trans*, or III) As yet undetermined interactions with other protein complexes to facilitate indirect binding to DNA.